



Fouille de données spatiales et modélisation de linéaires de paysages agricoles

Sébastien da Silva

► To cite this version:

Sébastien da Silva. Fouille de données spatiales et modélisation de linéaires de paysages agricoles. Informatique [cs]. Université de Lorraine, 2014. Français. tel-01101424

HAL Id: tel-01101424

<https://hal.inria.fr/tel-01101424>

Submitted on 8 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fouille de données spatiales et modélisation de linéaires de paysages agricoles

THÈSE

présentée et soutenue publiquement le 11 Septembre 2014

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention informatique)

par

Sébastien Da Silva

Composition du jury

<i>Rapporteurs :</i>	Marie-Odile CORDIER Didier JOSSELIN	Professeur, Université de Rennes Directeur de Recherche CNRS, Avignon
<i>Examineurs :</i>	Katarzyna ADAMCZYK Isabelle DEBLED-RENNESON Alexandre JOANNON Amedeo NAPOLI	Ingénieur de Recherche, INRA Jouy-en-Josas Professeur, Université de Nancy Chargé de Recherche INRA, Rennes Directeur de Recherche CNRS, Nancy
<i>Directrices :</i>	Florence LE BER Claire LAVIGNE	Directrice de la Recherche, ENGEEES Strasbourg Directrice de Recherche, INRA-PACA Avignon

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Mis en page avec la classe thesul.

Remerciements

*Je dédie cette thèse
à mes anges gardiens*

Sommaire

Chapitre 1 Introduction	xv
--------------------------------	-----------

Partie I Contexte

Chapitre 2 Problématique	3
---------------------------------	----------

Chapitre 3 État de l'art	7
---------------------------------	----------

3.1 Méthodes statistiques et géométriques pour l'analyse et la génération d'information spatiale	8
3.1.1 Méthodes de caractérisation de données spatiales	8
3.1.2 Prise en compte de la proximité spatiale	11
Rappel	11
Pseudo-distance	12
3.1.3 Méthodes de pavage	13
3.2 Courbe remplissant l'espace	14
3.2.1 Présentation	15
3.2.2 Courbe de Hilbert	17
3.2.3 Chemin de Hilbert adaptatif	17
Présentation	18
3.3 Méthodes de Markov	20

3.3.1	Rappels	20
3.3.2	Utilisation des modèles de Markov	21

Partie II Pré-traitements et fouille de données spatiales

Chapitre 4	Données et pré-traitements	27
4.1	Présentation des données	28
4.1.1	Description des zones d'études	28
4.1.2	Production des données	28
4.2	Préparation des données	29
4.2.1	Découpe en segments	29
4.2.2	Découpe en cellules	29
4.2.3	Classification des cellules	31
4.3	Analyse des données	32
4.3.1	Dans leur globalité	32
4.3.2	Par classe	34

Chapitre 5	Caractérisation des structures spatiales de segments par l'étude de leurs voisinages	37
5.1	Études de voisinages et densité de segments dans l'espace	38
5.1.1	Outils développés	38
	Distance entre segments	38
	Voisinage d'un segment	39
	Angle entre segments	40
	Densité relative de différents types de segments dans le voisinage d'un segment	41
	Densité relative, à l'échelle de la cellule, de différents types de segments dans le voisinage d'un segment	42

5.1.2	Tests	43
	Test des facteurs qui impactent les densités relatives	43
	Test de significativité des densités relatives observées	43
5.2	Densité relative au voisinage des segments de haies	44
5.2.1	Le plus proche voisin d'un segment de haie	44
5.2.2	Densité relative de segments de haies, routes et canaux dans les voisinages de segments de haies	44
	Distributions des voisins sur deux cellules typiques	44
	Synthèse en considérant l'ensemble des cellules	47
	Effet de l'orientation relative des voisins	48
	Comportement différent des segments de haie de type HP et HV	48
	Hétérogénéité au sein des paysages : l'effet "classe" . .	53

Chapitre 6 Apprentissage sur les structures spatiales pour les gé- rer 55

6.1	Linéarisation de l'information spatiale avec le chemin de Hilbert adaptatif	56
6.1.1	Définition de case et profondeur de découpe dans le chemin de Hilbert adaptatif	56
6.1.2	Définition de temps de parcours et temps d'attente dans le <i>CHA</i>	57
6.2	Apprentissage par chaînes de Markov	59
6.2.1	Markov sur le chemin de Hilbert adaptatif	59
6.2.2	Distance entre matrices de transition	59
6.3	Création du chemin de Hilbert adaptatif pour les données A et B .	60
6.3.1	Caractérisation des chemins de Hilbert adaptatifs par les pro- fondeurs de découpe	60
6.3.2	Caractérisation des chemins de Hilbert adaptatifs par les temps d'attente	63
6.4	Utilisation des chaînes de Markov sur les informations linéarisées pour les données A et B	65
6.4.1	Calcul des matrices de transition	65
6.4.2	Classification des matrices de transition	67

Variable <i>Longueur du segment</i>	67
Variable <i>Angle du segment</i>	68

Partie III Simulation et évaluation

Chapitre 7 Génération de structures de segments dans l'espace	73
7.1 Stratégie	74
Données	74
Étape 1 : Création d'une cellule vide	74
Étape 2 : Simulation du positionnement du milieu des segments	75
Étape 3 : Simulations indépendantes pour les variables <i>Longueur du segment</i> et <i>Angle du Segment</i>	76
Étape 4 : Attribution de valeurs pour les variables <i>Longueur du segment</i> et <i>Angle du segment</i>	77
Étape 5 : Simulation de la cellule suivante	79
7.2 Cellules simulées par génération de structures de segments dans l'espace - Comparaison au réel	80
7.2.1 Indicateurs de statistiques descriptives	81
Indicateur 1 : Nombre de segments de haies par cellules . . .	81
pour le paysage A :	81
pour le paysage B :	81
Indicateur 2 : Proportion des segments de type HV et HP par cellules	84
De type HV dans le paysage A :	84
De type HV dans le paysage B :	84
De type HP dans le paysage A :	87
De type HP dans le paysage B :	87
Indicateur 3 : Longueur des segments de haies par cellules .	90
7.2.2 Chemins de Hilbert adaptatifs	92

7.3	Génération améliorée - Ajustement à partir des connaissances du domaine	94
-----	---	----

Partie IV Conclusion

Chapitre 8 Conclusion et perspectives	99
---------------------------------------	----

Partie V Bibliographie

Bibliographie	105
Annexe	115
Annexe A Manuel utilisateur de l’Outil	115
A.1 Interface et utilisation	115
A.1.1 Partie Prétraitement	117
Découpe	119
Nombre d’éléments par cellule	120
Résumé des informations	121
A.1.2 Partie Statistiques	122
A.1.3 Partie Apprentissage	124
Création du fichier résumé de classe	126
Classification des valeurs	126
A.1.4 Partie Simulation	128
Création de la liste des centres de cases	128
Création de la liste des centres de cases classés	129
A.1.5 Dessin des barycentres simulés	132
A.1.6 Dessins des segments de haies simulés	133

Table des figures

1.1	Un aperçu des étapes qui composent le processus d'Exploration de Données [30]	xvii
1.2	Un aperçu des étapes qui composent le cheminement du travail de recherche durant la thèse	xxii
3.1	Exemples pour la comparaison visuelle des différentes distances de \mathbb{R}^2 .	12
3.2	Premières itérations pour la construction d'une courbe de Peano	15
3.3	Les trois premières itérations pour la construction d'une courbe remplissant l'espace à partir de trois motifs différents	16
3.4	Motifs pour la construction d'une courbe remplissant l'espace	17
3.5	Exemples de cases créées par division d'une cellule carrée	18
3.6	Exemple de construction de chemin de Hilbert adaptatif dans une cellule, d'après un nuage de point.	19
4.1	Plan du chapitre 4 : Création de données pré-traitées et transformées .	27
4.2	Géographie des cellules pour les deux jeux de données	31
4.3	Représentation d'une zone cellulaire, la cellule cible au centre (en gris foncé), les voisines autour (en bleu clair)	32
4.4	Histogrammes circulaires des angles pour les deux jeux de données . . .	33
4.5	Visualisation des classes obtenues par classification hiérarchique pour les deux jeux de données	34
5.1	Plan du chapitre 5 : Caractériser les structures spatiales par l'étude de leurs voisinages	37
5.2	Schéma explicatif pour le calcul de la distance entre deux segments . .	40
5.3	Visualisation du voisinage à une distance b d'un segment S de longueur L , pour la distance DiSt	40

5.4	Proportion de chaque type de voisin le plus proche pour chaque classe de cellule. Le résultat est présenté pour chaque paysage (à gauche : basse vallée de la Durance, à droite : Bretagne), pour chaque type de haie (1) : haies HV, 2) haies HP) et pour les orientations parallèles et perpendiculaires. Le type de voisin est route (noir), haie (gris clair) ou canal (gris foncé)	45
5.5	Deux cellules typiques et leur zone cellulaire dans la basse vallée de la Durance ($A-5_6$ à gauche) et en Bretagne ($B-9_2$ à droite). Le carré central représente la cellule de dimension $1100m \times 1100m$. Les lignes vertes représentent les haies, les noirs représentent les routes et les bleus représentent les canaux.	46
5.6	Densité relative $D_r^I(\mathcal{C}, b, C(\theta))$ de chaque type d'éléments (haies, routes, canaux) dans un voisinage croissant (de $20m$ à $500m$) autour des segments de haies pour deux cellules typiques ($A-5_6$ pour la basse vallée de la Durance et $B-9_2$ pour la Bretagne). $D_r^I(\mathcal{C}, b, C(\theta))$ est donné pour les deux orientations relatives (parallèle et perpendiculaire) et dans le voisinage de tous les segments de haies (à gauche), des haies HP (au centre) ou des haies HV (à droite).	47
5.7	Densité relative normalisée de chaque type de segments (en haut : routes, au milieu : haies, en bas : canaux) à des distances croissantes des segments de haies dans chaque paysage (Bretagne à gauche et basse vallée de la Durance à droite). Les traits pleins représentent les segments de haies HP, les traits en pointillés représentent les segments de haies HV, les ronds noirs sont présents sur les courbes concernant l'orientation parallèle du voisinage et les courbes sans ronds noirs concernent l'orientation perpendiculaire du voisinage	49
5.8	Densité relative normalisée de voisins du type routes ou haies, pour deux tailles de voisinages (20 m et 100 m de distance). Les résultats sont présentés pour tous les segments de haies du paysage de Bretagne (en haut), et pour les segments de haies HV (au milieu) ou de haies HP (en bas) pour le paysage de la basse vallée de la Durance. Les lignes grises correspondent aux tendances du changement pour les voisins avec l'orientation relative perpendiculaire et les lignes noires pour les voisins avec une orientation relative parallèle. Chaque ligne correspond à une classe avec son numéro correspondant indiqué sur la gauche.	52

6.1	Plan du chapitre 6 : Caractériser les structures spatiales pour les générer	55
6.2	Forme additive du temps de parcours	58
6.3	Proportion de la distribution de la variable <i>Profondeur de Découpe</i> pour une cellule moyenne, selon chaque classe, pour les données A. La profondeur de découpe est en abscisse. L'ordonnée est exprimée en pourcentage, par rapport à l'ensemble des valeurs prises par la variable <i>Profondeur de Découpe</i> sur l'ensemble de la cellule moyenne.	61
6.4	Proportion de la distribution de la variable <i>Profondeur de Découpe</i> pour une cellule moyenne, selon chaque classe, pour les données B. La profondeur de découpe est en abscisse. L'ordonnée est exprimé en pourcentage, par rapport à l'ensemble des valeurs prises par la variable <i>Profondeur de Découpe</i> sur l'ensemble de la cellule moyenne.	62
6.5	Représentation des classes d'angle pour les données.	66
6.6	Représentation graphique du résultat de la classification des matrices de transition suivant les deux variables, pour les données A. Chaque couleur correspond à une classe.	67
6.7	Représentation graphique du résultat de la classification des matrices de transition suivant les deux variables, pour les données B. Chaque couleur correspond à une classe.	69
7.1	Plan du chapitre 7 : Générer des segments de haies dans le plan	73
7.2	Exemple de probabilité de transition d'une variable <i>Angle</i> pour la classe $[\frac{\pi}{3}; \frac{\pi}{2}]$	79
7.3	Exemples de cellules simulées	80
7.4	Nombre de segments de haies par cellule pour les données A. En abscisse : le code des cellules du paysage A et la classe dans laquelle elles sont. En ordonnée : Le nombre de segments de haies.	82
7.5	Nombre de segments de haies par cellule pour les données B. En abscisse : le code des cellules du paysage B et la classe dans laquelle elles sont. En ordonnée : Le nombre de segments de haies.	83
7.6	Proportion des segments de type HV pour un échantillon de cellules issues du paysage A. En abscisse : le code des cellules du paysage A et la classe dans laquelle elles sont.	85
7.7	Proportion des segments de type HV pour un échantillon de cellules issues du paysage B. En abscisse : le code des cellules du paysage B et la classe dans laquelle elles sont.	86

7.8	Proportion des segments de type HP pour un échantillon de cellules issues du paysage A. En abscisse : le code des cellules du paysage A et la classe dans laquelle elles sont.	88
7.9	Proportion des segments de type HP pour un échantillon de cellules issues du paysage B. En abscisse : le code des cellules du paysage B et la classe dans laquelle elles sont.	89
7.10	Pour chaque cellule, proportion de chaque classe de longueur, en moyenne pour les 100 cellules générées (cercle noir), et pour la cellule réelle (cercle rouge). Les cellules sont issues du paysage A.	91
7.11	Distribution de la variable <i>Profondeur de Découpe</i> pour 100 cellules simulées (en noir) et pour la cellule réelle correspondante (en rouge). En ordonnées : la proportion de chaque valeur de profondeur de découpe sur chaque cellule.	93
7.12	Processus ajusté à partir des connaissances du domaine avec retour sur trace.	95
7.13	Résultat d'une cellule simulée (à gauche) pour la cellule réelle issue des données A (à droite) par la génération ajustée avec connaissance du domaine.	95
7.14	Résultat d'une cellule simulée (à gauche) pour la cellule réelle issue des données B (à droite) par la génération ajustée avec connaissance du domaine.	96
A.1	Présentation de l'interface et des différents éléments qui la constituent .	116
A.2	Présentation de l'onglet extraction	117
A.3	Présentation de l'onglet découpe	118
A.4	Présentation de l'onglet de calcul de distance entre les éléments	122
A.5	Présentation de l'onglet distance	123
A.6	Présentation de l'onglet densité	124
A.7	Présentation de l'onglet Hilbert adaptatif	125
A.8	Présentation de l'onglet classement des variables	126
A.9	Présentation de l'onglet création des matrices de transition	127
A.10	Présentation de l'onglet création cellule vide	128
A.11	Les trois premières itérations pour la construction d'un chemin de Hilbert	129
A.12	Présentation de l'onglet Simulation des points pour une cellule	130
A.13	Présentation de l'onglet de simulation pour un paysage	132
A.14	Présentation de l'onglet dessin pour une cellule	132